

# METADATA: Enumeration Areas Master File

Dataset: enumeration\_areas\_master.csv

## OVERVIEW

- **Purpose:** Master sampling frame of enumeration areas for survey sampling
- **Coverage:** Six SADC countries (Angola, Botswana, Comoros, DRC, Eswatini, Lesotho)
- **Base Year:** 2022 Population and Housing Census
- **Update Cycle:** Annual updates for new developments and boundary changes
- **Total Units:** 5,000 enumeration areas
- **Geographic Level:** Enumeration Area (smallest census geographic unit)

## SAMPLING FRAME CHARACTERISTICS

- **Frame Type:** Area frame based on census enumeration areas
- **Coverage Rate:** Estimated 98% of target population
- **Exclusions:** Military bases, diplomatic compounds, institutional populations
- **Update Sources:** Census cartography, satellite imagery, administrative records

## VARIABLE DEFINITIONS

### IDENTIFICATION AND LOCATION

- **ea\_id** [String]: Unique EA identifier (format: EA#####)
- **country** [String]: SADC member country
- **province\_code** [String]: Province identifier (P1-P8)
- **district\_code** [String]: District identifier (D1-D25)
- **stratum** [String]: Stratification variable (country\_urban/rural)
- **urban\_rural** [String]: Urban/Rural classification based on census definition
- **ea\_type** [String]: Detailed EA classification
  - Urban: Formal / Informal / Mixed
  - Rural: Village / Settlement / Farm

### GEOGRAPHIC COORDINATES

- **longitude** [Numeric]: Longitude of EA centroid (decimal degrees)

- **latitude** [Numeric]: Latitude of EA centroid (decimal degrees)
- **elevation** [Integer]: Elevation in meters above sea level

## SIZE MEASURES

- **total\_households** [Integer]: Total households from census
- **total\_population** [Integer]: Total population from census
- **households\_listed** [Integer]: Households from recent listing operation
- **last\_census\_year** [Integer]: Year of last census (2022)
- **listing\_date** [Date]: Date of most recent listing update

## ACCESSIBILITY INDICATORS

- **road\_distance\_km** [Numeric]: Distance to nearest all-weather road (km)
- **health\_facility\_km** [Numeric]: Distance to nearest health facility (km)
- **school\_km** [Numeric]: Distance to nearest primary school (km)
- **market\_km** [Numeric]: Distance to nearest market/trading center (km)
- **mobile\_coverage** [String]: Mobile network coverage
  - None: No mobile coverage
  - 2G: Basic mobile service
  - 3G: Mobile data service
  - 4G: High-speed data service
- **accessibility** [String]: Overall accessibility classification
  - Easy: Accessible year-round by vehicle
  - Moderate: Seasonal access limitations
  - Difficult: Requires 4WD or walking
  - Very Difficult: Remote, requires special arrangements

## SAMPLING INFORMATION

- **selection\_probability** [Numeric]: Probability of selection in PPS sampling
- **psu\_weight** [Numeric]: Base weight for primary sampling unit
- **quality\_flag** [Binary]: Data quality indicator
  - 0: No quality issues
  - 1: Quality issues noted (check notes)

## STRATIFICATION RULES

1. **Primary Stratification:** Country
2. **Secondary Stratification:** Urban/Rural
3. **Implicit Stratification:** Geographic sorting within strata
4. **Size Stratification:** EAs grouped by household count for PPS

## SIZE CLASSIFICATIONS

- **Small EA:** <50 households
- **Medium EA:** 50-150 households
- **Large EA:** 151-300 households
- **Very Large EA:** >300 households (consider splitting)

## QUALITY INDICATORS

### Coverage Assessment

- **Undercoverage Sources:**
  - New developments since census
  - Informal settlements growth
  - Seasonal migration areas
- **Overcoverage Sources:**
  - Demolished areas
  - Converted residential areas
  - Depopulated rural areas

### Quality Flags Interpretation

- Flag 0: Standard quality, use normally
- Flag 1: Review before selection
  - Boundary changes detected
  - Significant population change
  - Listing discrepancies >20%
  - Missing geographic coordinates

## UPDATE PROCEDURES

### 1. Annual Updates:

- New residential developments added
- Boundary adjustments processed
- Population estimates updated

### 2. Listing Operations:

- Selected EAs listed before survey
- Dwelling units enumerated and mapped
- GPS coordinates verified

### 3. Quality Control:

- 10% of EAs verified independently
- Satellite imagery comparison
- Administrative record matching

## PPS SELECTION PROCEDURES

For PPS selection:

1. Sort EAs by stratum and geographic order
2. Calculate cumulative measure of size (households)
3. Generate systematic random start
4. Select EAs at regular intervals
5. Document selection probabilities

## MISSING DATA PATTERNS

- **Coordinates:** 2% missing due to security/access issues
- **Distance measures:** Complete for all EAs
- **Population counts:** Updated from projections where needed
- **Mobile coverage:** Based on 2023 network maps

## LINKAGE INFORMATION

- Links to **household\_survey\_main\_2024.csv** via **ea\_id**
- Links to **auxiliary\_census\_2022.csv** via **district\_code**
- Links to **small\_area\_indicators.csv** via **district\_code**

## USAGE GUIDELINES

1. **Sampling:** Use selection\_probability for PPS designs
2. **Weighting:** Start with psu\_weight for design weights
3. **Stratification:** Respect stratum boundaries in selection
4. **Quality:** Check quality\_flag before selection
5. **Updates:** Verify listing\_date for currency

## GEOGRAPHIC PROJECTION

- **Coordinate System:** WGS 84 (EPSG:4326)
- **Datum:** World Geodetic System 1984
- **Units:** Decimal degrees
- **Precision:** 6 decimal places

## TEMPORAL COVERAGE

- **Census Base:** 2022 Population and Housing Census
- **Frame Creation:** January 2023
- **Last Major Update:** July 2023
- **Next Scheduled Update:** January 2025

## DATA LIMITATIONS

1. **Geographic Coverage:**
  - Some border areas have restricted access
  - Island territories may have outdated boundaries
2. **Population Dynamics:**
  - Rapid urbanization may outpace updates
  - Seasonal population movements not fully captured
3. **Infrastructure Changes:**
  - New roads and facilities updated annually
  - Mobile coverage data from network operator reports

## FILE SPECIFICATIONS

- **Format:** CSV (Comma Separated Values)
- **Encoding:** UTF-8

- **Size:** Approximately 5,000 records
- **Headers:** First row contains variable names
- **Delimiter:** Comma (,)
- **Text Qualifier:** Double quotes for text fields with commas

## VERSION CONTROL

- **Version:** 2.0
- **Release Date:** 2024-01-15
- **Previous Version:** 1.0 (2023-01-15)
- **Change Log:** Available in EA\_Changes\_Log.txt

## QUALITY ASSURANCE

- **Validation Rate:** 100% automated checks
- **Manual Review:** 10% random sample
- **GPS Verification:** Field verified for sampled EAs
- **Consistency Checks:** Cross-validated with census data

## CONTACT INFORMATION

- **Data Manager:** SADC Statistical Unit
- **Email:** [sampling.frame@sadc-stats.org](mailto:sampling.frame@sadc-stats.org)
- **Phone:** +267 XXX XXXX
- **Address:** SADC Secretariat, Gaborone, Botswana

## CITATION

SADC Statistical Unit (2024). Enumeration Areas Master Sampling Frame, Version 2.0.  
Southern African Development Community, Gaborone, Botswana.  
Based on 2022 Population and Housing Census.